CST1204: Introduction to Databases

Jun Shan
Adjunct Professor
City Tech, CUNY

Agenda

- Course Introduction
- Introduction to database concepts (Part 1 of 2)
  - Relational databases
  - Entity, Attribute, and Relationship
  - Functional Dependence
  - Primary keys
  - Database Design

Course Introduction

- Course syllabus
  - Grading requirements
- Environment setup

Database Concepts (Part 1 of 2)

- Relational databases
- Entity, Attribute, and Relationship
- Functional Dependence
- Primary keys
- Database Design

Relational Model

- **Database: a structure that contains different categories of information and the relationships between these categories. (Ch1. Pg 1)**
- **Relational Model by Dr. Edgar F. Codd (1970):**
  - **1. Even the most complex datasets can be represented by two types of objects, *Entity* and *Relationship*.**
  - **2. Both Entity and Relationship can be logically represented in computer systems.**

Relational Database

- A collection of tables. Formally, these tables are called relations, and this is how this type of database gets its name. (Ch 2, Pg 22)
- Relational database products: Oracle, SQL Server, MySQL, PostgreSQL
- Relational database examples:
  - - Textbook example: Ch 2, Pg 22
  - - Oracle Live SQL example

Entity, Attribute, and Relationship

- **Entity: a noun; it is a person, place, thing, or event.**
- **Attribute: a property of an entity.**
- **Relationship: the association between entities.**
  - **One-to-many**
  - **One-to-one: Usually can be considered as Attribute.**
  - **Many-to-many**

Database Representation

- Always two dimensional tables in RDB
- Entity/Attribute: One entity per table, and attribute are columns
- Relationship: the association between entities.
  - Two direction linkage
  - Repeating group
- Shorthand representation

Functional Dependency

- In a relational database, column B is functionally dependent on another column (or a collection of columns), A, if at any point in time a value for A determines a single value for B. (Ch 2, Pg 28)
- Functional Determination
- The only way to determine functional dependencies is to examine the user's policies. This is called "data discovery".

Primary Key

- **The unique identifier for a table (Ch 2, Pg 30)**
  - **Property 1: All columns in the table are functionally dependent on A.**
  - **Property 2: No subcollection of the columns in A (assuming A is a collection of columns and not just a single column) also has property 1.**

Database Design

- Design method
- Database design requirements
- Database design process example

Note: CST3504 "Database Design" will cover processes and principles of database design in detail.

In Class Review

- **Chapter 2 Review Questions 1-11**
- **Chapter 2 exercise questions for TAL Distributors (1, 2, 3, and 5)**

Homework

- Design a database for Colonial Adventure Tours using shorthand representation (review the tables in Chapter 1)
  - Write down whether each table is an entity or relationship
  - Write down primary keys
  - Write down functional dependencies

Agenda

- Introduction to database concepts (Part 2 of 2)
  - Normalization (first, second and third normal form)
  - Diagrams for database design
  - Database Design (Review)

Review of Last Lecture

- Relational databases
- Entity, Attribute, and Relationship
- Functional Dependence
- Primary keys
- Database Design

Normalization

- **Normalization: The process to identify the existence of potential problems, such as data duplication and redundancy, and implement ways to correct these problems. The goal of this process is to allow you to take a table or collection of tables and produce a new collection of tables that represents the same information but is free of problems. (Ch 2, Pg 40)**
- **Normal Form (NF): A table in a particular normal form possesses a certain desirable collection of properties. (Ch 2, Pg 40)**

First Normal Form

- **A table (relation) is in first normal form (1NF) when it does not contain a repeating group. (Ch 2, Pg 40)**

Second Normal Form

- A table (relation) is in second normal form (2NF) when it is in first normal form and no nonkey column (that is, a column that is not part of the primary key) is dependent on only a portion of the primary key. (Ch 2, Pg 44)

Third Normal Form

- **A table is in third normal form (3NF) when it is in second normal form and the only determinants it contains are candidate keys. (Ch 2, Pg 48)**

Note: The form as stated above is in fact Boyce-Codd Normal Form (BCNF), an enhanced version of the original 3NF.

Review Previous Homework

Verify your design against 1st, 2nd, and 3rd normal forms.

Denormalization

- Normalization may cause performance issue
- Denormalization to put groups of attributes as one block
    - e.g., State in an address record

Diagrams for database design

- **Entity-Relationship Diagram (ERD)**
- **Arrowed diagram**
- **Crowfoot diagram**
- **Original diamond shape diagram**

In Class Review

- Chapter 2 Review Questions 12-17
- Chapter 2 exercise questions for TAL Distributors

Optional

- Design and normalization of Solmaris Condominium Group

Homework

- **Chapter 2 exercise questions for Colonial Adventure Tours**